Research on Hybrid Recommendation Algorithm of Mother and Child Information Based on Tagging System

Lin Zhong

Shanghai University, Shanghai 200436, China.

Abstract: The development of mobile news and information makes the access to information resources more convenient, however, the network of Shanghai's information also brings some trouble to users to read valuable information efficiently. Taking the mother and baby information project on the home page of MeetYou APP as an example, by marking the information information and dynamically marking the 50 million users in the APP. Using the recommendation algorithm based on weight, location and collaborative filtering, combined with the calculation of inverse word frequency and cosine similarity between tags, a hybrid tagging recommendation algorithm is proposed. Tested in a practical project, the algorithm can effectively improve the efficiency of personalized recommendation and provide reference value for optimizing the information content recommendation system.

Keywords: Tagging System; Text Tagging; User Tagging; Recommendation Algorithm; Collaborative Filtering; User Behavior Log; Personalized Recommendation

1. Research Background

With the development of technology, the news and information industry has experienced the traditional media era and the online media era, coming to the era of intelligent recommended self-media; the publishers, dissemination channels and recommendation mechanisms of news and information have undergone great changes. In the era of information overload, text, audio, video, images, social networks and other kinds of information on the Internet are growing in an explosive situation, enriching people's daily life and study and work content. Personalized recommendation is indeed an important mechanism to save time for users. There are more and more information applications on the market, such as Today's Headlines, Watch Recommend, Instant, etc. However, in the case of MeetYou, which has a specific recommendation, it is a good idea to use it. However, in an app with a specific female user demographic like MeetYou, the large amount of knowledge content makes it more difficult for users to find the content they like. Since mother and baby knowledge is rich in variety, large in quantity, short in length, and coherent and sequential, traditional single recommendation algorithm cannot solve the problem of accuracy of knowledge recommendation in a targeted way. For example, Today's Headline uses recommendation algorithm to recommend content based on users' past reading behaviors as user interests, which is mostly repetitive and redundant; One Point adds the rules of subscription on the basis of algorithm to improve the accuracy of recommendation. In addition to using algorithms, See Recommend also introduces manual screening into the chain of personalized recommendation. In this paper, we conduct deep data mining based on user preferences and knowledge tags, and use collaborative filtering and label-based CRT hybrid recommendation algorithm to recommend suitable mother and baby knowledge for users, so as to improve personalized recommendation efficiency and provide a reference method for optimizing mother and baby knowledge recommendation system.

2. Labeling system

In this paper, the tagging system includes two parts: text tagging and user tagging. Text tagging is to identify and tag information, i.e., knowledge itself, to extract text features; while user tagging is a dynamic tagging of user portrait features constructed through data mining and mathematical modeling. The tagging system is to build up the connection between users and information through tags, so as to achieve better content-based user personalized recommendation, and the tagging system is shown in Figure 1 below.

2.1 Text labels

In this paper, the information is better organized and classified through tags, which are designed into three levels according to the characteristics of the actual project: industry classification, topic tags, and keyword tags, which belong to a structured tagging system, and the industry classification is the parent of the topic tags. An example of classification is shown in Figure 2 below.

Figure 2-1

Each text enters the system and needs to go through the text tagging process. Firstly, it is tagged with keyword tags, then it is tagged with subject tags, and finally, according to the mapping relationship between primary and secondary categories, the corresponding industry classification tags are obtained.

2.1.1 The processing flow of keyword tags is shown in Figure 1. Next, we will describe the implementation of each step in detail.

(1) First of all, we need to classify and annotate the text, and extract the feature keywords. This part of the work is mainly done by using HanLP tools.

(2) Calculate keyword weights

The model designed in this paper for calculating keyword weights considers three main aspects.

A.The tagging of tags, consider the strong subject matter, high interest point of the tagging weight is higher, the tagging of keyword tags is divided into: strong (1.0). General (0.6), weak (0.25), TWx in the formula: x tagging weight of tagging.

B.Label appearance position, consider the more important the location of the label appears, the higher the weight, appearing position: title (1.2), content head (0.8), content tail (0.6), content middle (0.4) PWxi in the formula: x the weight of the i-th appearance position of the label.

C.The number of times the label appears, consider the more times the label appears, the higher the weight, but the growth rate will become slower, the formula TFx: x number of times the label appears, the formula$\frac{1.0}{i}$: the i-th appearance of the label weighted amount;

Finally, based on the above three aspects to obtain the weighting formula for keyword tag x is as follows.

$$f(x) = TW_w \sum_{i}^{TF_x} \frac{PW_{xi}}{i} \quad (1)$$

2.1.2 The theme labeling model training idea is as follows.

A. Let there be N topic tags: {T_1, T_2,..., T_N}; the sum of the topic weight distribution of each keyword tag is agreed to be 1.0, and the initialized weight of each topic is the same; let the topic weight distribution of keyword tag k be {TW_1, TW_2,...,TW_N}, according to the above agreement then We have TW_1= TW_2=...= TW_N=1.0/N.

B. So how do we get the actual topic weight distribution for keyword tag k? Suppose that at some moment the training model flows through a piece of data labeled with topic tag i with keyword tag k. Then the increment of TW_i for keyword tag k will be contributed by tags other than i (1.0-TW_i) , let TW_i = TW_i + α (1.0-TW_i) , α is an empirical constant between 0 and 1, and 1.0/N is taken here; since 1.0-TW_i = TW_ 1+ TW_2+...+ TW_N-TW_i, then: TW_j = TW_j - TW_j/N, j ∈ {1,2,... ,N} and j ≠ i.

When the training data for keyword tag k is sufficient, the distribution of topic weights for keyword tag k will also tend to converge

The keyword labels of all training data are iterated once to complete the learning calculation of the topic label model (the set of keyword label topic distribution weights) {keyword label 1:{topic 1:weight 1,...},...}.

C. The training process of the keyword tag-based topic tagging model is as follows.

Figure 2-2

2.2 User tags

User tagging is a user portrait feature constructed through data mining and mathematical modeling. By tagging to better mine and locate user interests, the collection of all features of users becomes a tagging system. The user tags designed in this paper include two parts: user keyword tags and user topic tags. The following details will introduce the initialization and dynamic update process method of user tags.

2.2.1 User theme tag initialization

The initialization of the theme tags is assigned according to the different identities of female users. There are four different identities in the system, which are initialized according to the following theme tags, and then wait for the user to read the behavior log for dynamic updates.

Menstruation		Pregnancy		Preparedness		Hot Mom							
						0-3M		3-6M		6-12M		12M+	
5	Stars	32	Pregnancy	31	Preparedness	33	Childcare	33	Childcare	25	Emotions	25	Emotions
20	Skin Care	25	Emotions	25	Emotions	32	Pregnancy	25	Emotions	33	Childcare	33	Childcare
36	Costumes	31	Preparedness	32	Pregnancy	25	Emotions	5	Stars	5	Stars	5	Stars
25	Emotions	33	Childcare	5	Stars	5	Stars	32	Pregnancy	32	Pregnancy	32	Pregnancy
54	Society	54	Society	6	Funny	6	Funny	36	Costumes	36	Costumes	6	Funny
32	Pregnancy	6	Funny	33	Childcare	36	Costumes	6	Funny	6	Funny	36	Costumes
6	Funny	5	Stars	30	Menstruation	31	Preparedness	31	Preparedness	18	Gourmet	30	Menstruation
33	Childcare	36	Costumes	24	Sexual	30	Menstruation	18	Gourmet	9	Home design	18	Gourmet
31	Preparedness	9	Home design										
30	Menstruation												

Figure 2-3

2.2.2 User theme tag update process

Based on the user reading behavior log to update and use the Hbase theme tagging model, mainly read the real-time data of user exposure and click, sorted by time ascending order, give positive and negative feedback and overwrite the original Hbase data. The main model ideas are as follows.

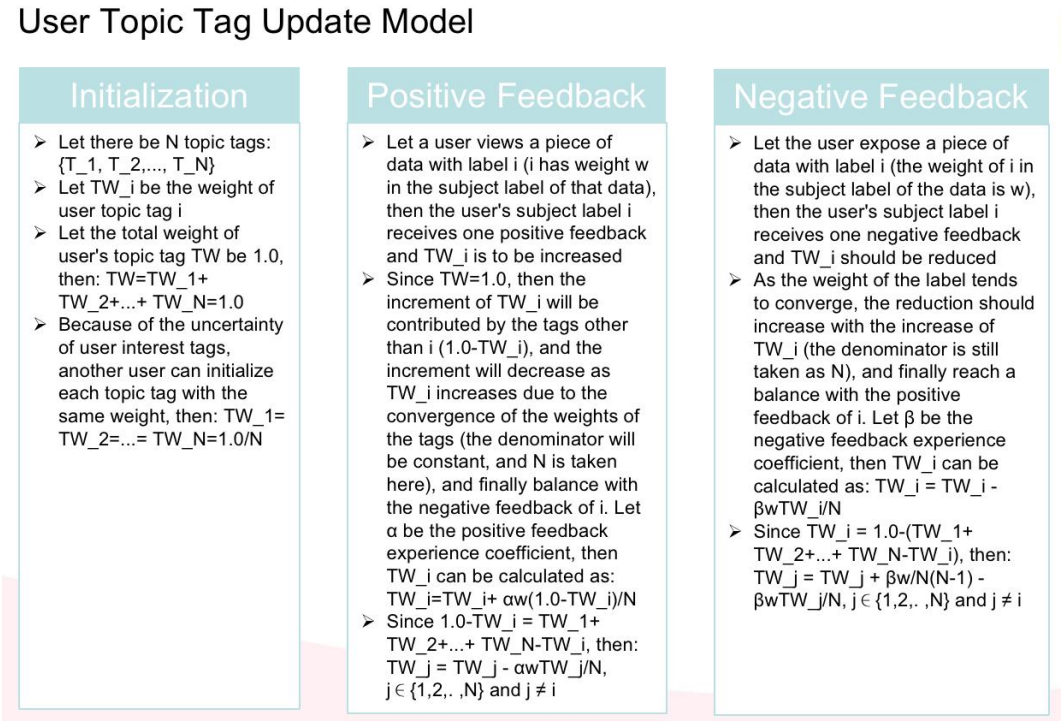

Figure 2-4

User keyword tag update process ideas as above, this article skipped.

3. Personalized information recommendation system

The experimental data in this paper comes from the data of the largest mother and baby women's platform in China, MeetYou APP, in which tens of thousands of pieces of information are added to the mother and baby information library every day. Users can get free maternal and infant information knowledge services on the home page of MeetYou APP. 8 million daily users within this APP generate a huge amount of user data. These data include user identity, information record, information name, industry classification, topic tag, keyword tag, friend relationship and other rich information. The personalized recommendation system is an information filtering system to help users reduce the time and energy wasted by browsing a large amount of invalid data. In this paper, we mainly use a mixture of various recommendation algorithms based on weight and location-based hybrid algorithms. The schematic diagram of the main algorithmic data pool is as follows.

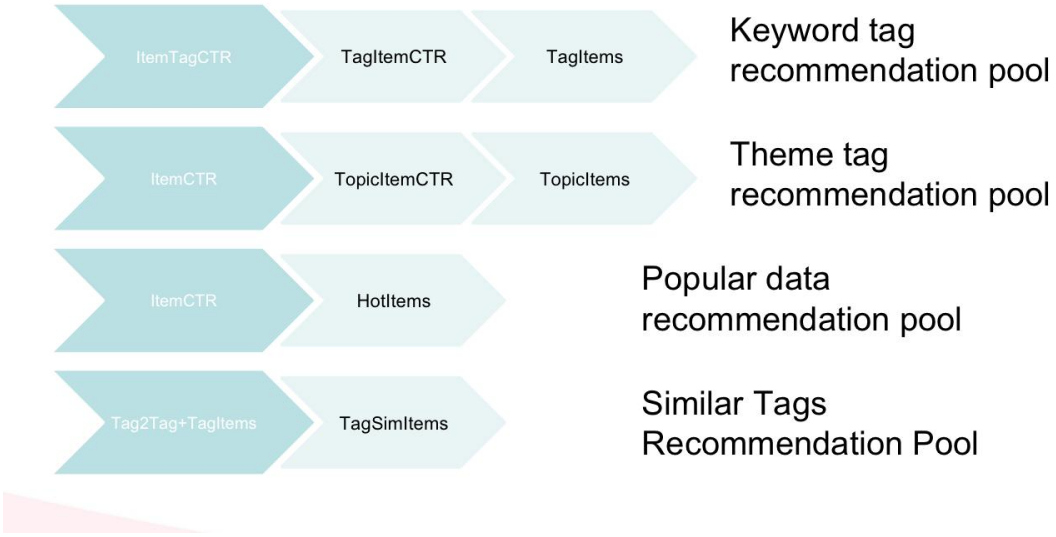

Figure 3-1

3.1 Weight-based recommendation logic

In the personalized information recommendation system, a weight-based hybrid recommendation algorithm is used, for example: according to the actual number of requests, various types of recommendation data are allocated, and the current allocation strategy is: new keyword tags account for 0.2, strong interest keyword tags (user interest > 0.2) account for 0.8; the insufficient part of the complementary strategy is, in order: strong interest keyword tag data, general interest keyword tag (user interest <0.2) data, topic tag (weight > average weight) data, similar tag recommendation data, and popular data; each keyword tag only takes at most one recommendation data each time, and each topic tag does not exceed 5 recommendation data, and after enough data are taken, the data are sorted by weight: keyword tag related recommendation data W=2.5*ItemTagCTR+4* UserTagW^2 (similar tag UserTagW is set to 0.15), subject tag recommendation data W=4*ItemCTR, popular recommendation data W=2.5*ItemCTR.

3.2 Location-based recommendation logic

In the personalized information recommendation system, a hybrid recommendation algorithm based on location is used, for example: exposure location and exposure data type template: ['keyword tag','keyword tag','new keyword tag ', 'keyword tag', ['theme tag', 'diversity tag'], the insufficient part of the complementary strategy in order: strong interest keyword tag data, new keyword tag, general interest keyword tag (user interest <0.2) data, topic tag (weight > average weight) data, similar tag recommendation data, popular data; each keyword tag only takes at most one recommendation data each time, and each topic

tag does not exceed 4 recommendation data; each recommendation pool data sorting: keyword tag part W=2.5* ItemTagCTR+4*UserTagW^2, other directly sorted by ItemCTR.

4. Conclusion

With the rapid development of mobile information business, the information knowledge base is getting richer and richer, and the differentiation of users' preferences is getting bigger and bigger. One of the difficulties in the promotion of mother and child knowledge business now is how to make accurate personalized recommendations to users from the huge information knowledge base conveniently and quickly. In this paper, we analyze the massive user reading behavior records saved in MeetYou APP, determine text tags and user dynamic tags through the tagging system, and then use a hybrid recommendation algorithm based on weight and location to make personalized information knowledge recommendations for users, which provides a new idea and reference to the recommendation algorithm in the information knowledge field.

References

[1] Huang, CL., Lu, YX., Research on hybrid music recommendation algorithm based on collaborative filtering and tagging[J]. Software Engineering,2021(4).01-03.

[2] Jia, YD., Liu, L., A study on identification and classification methods for Chinese non-functional requirement descriptions.Journal of Software, 2019, 30(10): 3115-3126.

[3] Zhuan, Y., Xiong, JH., Cheng, XQ., A character tagging recommendation method integrating personalization and diversity [J]. Journal of Chinese Information, 2017 (2): 154-162.